**Curved Stellar Paths and Proper Motion Determination**

George H. Kaplan
*U.S. Naval Observatory*

**Abstract** The publication of the Hipparcos catalog has drawn new attention to the problem of the determination of proper motions of stars that are components of binary or multiple systems. With only three years of observational data, it was recognized that unmodeled binary motion could affect the measured proper motions of many stars in the Hipparcos catalog. In fact, the problem occurs on many time scales and undoubtedly affects many catalogs. This note presents some analytical expressions for the effects of binary motion on proper motions when the orbital period is at least several times the span of observations.

**Key Words** astrometry, proper motions, positional astronomy, binary stars, catalogs

## 1. Introduction

The publication of the Hipparcos catalog has drawn new attention to the problem of the determination of proper motions of stars that are components of binary or multiple systems. With only about three years of observational data, it was recognized that unmodeled binary motion could affect the proper motions of many stars in the Hipparcos catalog. Most of the concern has been with binaries without known orbits that may have periods of several years to several decades. Yet almost half of all known binary systems have semimajor axes of 50 AU or more, implying, for solar-mass components, periods of a century or more. Such systems can create problems when observational catalogs made years or decades apart are combined (as in Tycho 2) in an effort to provide improved proper motions. Thus, the binary-motion problem occurs on many time scales and may affect, to some degree, a significant fraction of the data in compiled catalogs.

In this paper we look at a particularly pernicious piece of this problem: stellar paths that are almost, but not quite, linear. We provide a simplified model of astrometric observation analysis to quantify the problem. We then use the results to estimate the range of orbital semimajor axes that the observations are sensitive to, as a function of observational accuracy and time span, and the distance and mass of the system.

## 2. Simplified Development of Proper Motion Estimation

In this section we develop a simple model of how analysis of astrometric observations for stellar proper motion determination can become contaminated by an unmodeled acceleration. The object is to provide approximate expressions that will allow us to determine the order of magnitude of the effect as well as its qualitative nature.

The equation of motion of a body can be expressed as a Taylor series in vector form as

$$\mathbf{P}(t) = \mathbf{P}_0 + \mathbf{V}_0 t + \tfrac{1}{2}\mathbf{Z}t^2 + \cdots \tag{1}$$

where $\mathbf{P}_0$ and $\mathbf{V}_0$ are the body's position and velocity at time $t = 0$, and $\mathbf{Z}$ is the acceleration. time derivative. For a star where the acceleration is due to the gravitational attraction of a companion (either seen or unseen), $\mathbf{Z} = (GM/R^2)\hat{\mathbf{R}}$, where $G$ is the constant of gravitation and $M$ and $R$ are the mass and distance of the companion. The unit vector $\hat{\mathbf{R}}$ points toward the companion. If we assume a constant acceleration — i.e., truncate eq. (1) after the third term — we are limited to considering a small segment of the orbit. This is the problem we wish to investigate: the parabolic approximation constitutes a "weak curvature" case.

Let the star's motion, projected onto the plane of the sky, be $\mathbf{p}(t)$. The plane of the sky is orthogonal to the line of sight unit vector $\mathbf{n}$, so we have

$$\begin{aligned}\mathbf{p}(t) &= \mathbf{P}(t) - (\mathbf{P}(t)\cdot\mathbf{n})\mathbf{n} \\ &= (\mathbf{P}_0 + \mathbf{V}_0 t + \tfrac{1}{2}\mathbf{Z}t^2) - ((\mathbf{P}_0 + \mathbf{V}_0 t + \tfrac{1}{2}\mathbf{Z}t^2)\cdot\mathbf{n})\mathbf{n} \\ &= \mathbf{P}_0 + \mathbf{V}_0 t + \tfrac{1}{2}\mathbf{Z}t^2 - (\mathbf{P}_0\cdot\mathbf{n})\mathbf{n} - (\mathbf{V}_0\cdot\mathbf{n})t\mathbf{n} - \tfrac{1}{2}(\mathbf{Z}\cdot\mathbf{n})t^2\mathbf{n} \\ &= [\mathbf{P}_0 - (\mathbf{P}_0\cdot\mathbf{n})\mathbf{n}] + [\mathbf{V}_0 - (\mathbf{V}_0\cdot\mathbf{n})\mathbf{n}]t + \tfrac{1}{2}[\mathbf{Z} - (\mathbf{Z}\cdot\mathbf{n})\mathbf{n}]t^2\end{aligned} \tag{2}$$

That is, the star's motion in the plane of the sky can be represented as

$$\mathbf{p}(t) = \mathbf{p}_0 + \mathbf{v}_0 t + \tfrac{1}{2}\mathbf{z}t^2 \tag{3}$$

$$\begin{aligned} \text{where}\quad \mathbf{p}_0 \text{ is } & \mathbf{P}_0 \text{ projected onto the sky} = \mathbf{P}_0 - (\mathbf{P}_0 \cdot \mathbf{n})\mathbf{n} \\ \mathbf{v}_0 \text{ is } & \mathbf{V}_0 \text{ projected onto the sky} = \mathbf{V}_0 - (\mathbf{V}_0 \cdot \mathbf{n})\mathbf{n} \\ \mathbf{z} \text{ is } & \mathbf{Z} \text{ projected onto the sky} = \mathbf{Z} - (\mathbf{Z} \cdot \mathbf{n})\mathbf{n} \end{aligned}$$

The vectors $\mathbf{p}(t)$, $\mathbf{p}_0$, $\mathbf{v}_0$, and $\mathbf{z}$ have no component along the line of sight $\mathbf{n}$ and are therefore 2-vectors in a coordinate system on the plane of the sky.

In standard current practice, a star's motion is expressed as $\mathbf{p}'(t) = \mathbf{p}'_0 + \mathbf{v}'t$. Here, $\mathbf{v}'$ represents the star's proper motion as conventionally defined — assumed constant, hence without subscript. (In this development we are ignoring terms for the curvature of the sky and are assuming that aberration and parallax have already been removed from the data.) It is tempting to think of eq. (3) as a simple extension of the conventional expression, carried to higher order. But $\mathbf{v}_0$ in eq. (3) does not correspond to proper motion. For a gravitationally-bound binary, proper motion properly refers to the projection on the sky of the (constant) space velocity of the center of mass of the system. In eq. (3), $\mathbf{v}_0$ is the instantaneous linear component of the star's apparent motion, which is the sum of the proper motion of system plus the projected orbital velocity of the star (at $t$=0) around the center of mass. Determination of the proper motion of the system would require observations of the star spanning nearly an orbital period, or observations of both the star and its companion over a shorter period together with an estimate of their mass ratio.

But here we wish to consider a series of observations where the orbital motion of the star is not obvious. Consider an observation of the star's position, $\mathbf{p}_i$, taken at time $t_i$, with a measurement error $\mathbf{e}_i$ (all vectors are now in the plane of the sky). Then $\mathbf{p}_i = \mathbf{p}_0 + \mathbf{v}_0 t_i + \frac{1}{2}\mathbf{z}t_i^2 + \mathbf{e}_i$. But if we have no knowledge of the acceleration, we will model the star's motion in the conventional way as $\mathbf{p}'(t) = \mathbf{p}'_0 + \mathbf{v}'t$. The difference between the observation and this incomplete model of the star's motion at time $t_i$ is then

$$\mathbf{e}'_i = \mathbf{p}_0 + \mathbf{v}_0 t_i + \tfrac{1}{2}\mathbf{z}t_i^2 + \mathbf{e}_i - \mathbf{p}'_0 - \mathbf{v}'t_i \tag{4}$$

Compared to our incomplete model, the observation will appear to be in error by an amount $\mathbf{e}'_i$. This error is the sum of the random error of observation and the systematic error resulting from the use of the incorrect model. Despite the fact that it is not a purely stochastic quantity, it provides a basis for investigating what would happen if the incomplete model were fit to a set of $N$ two-dimensional observations using least squares. Let us take the common situation where the measurement errors $\mathbf{e}_i$ are assumed to all have approximately the same magnitude $\langle e \rangle$ and the observations are therefore all given unit weight. In such a case, the quantity that would be minimized is the sum of $\mathbf{e}'^2_i$ over all observations. The method would determine the position $\mathbf{p}'_0$ (at time $t$=0) and proper motion $\mathbf{v}'$ that minimize $\sum \mathbf{e}'^2_i$. The quantities $\mathbf{p}'_0$ and $\mathbf{v}'$ are not particularly interesting in themselves but the *differences* between these quantities and the corresponding ones from eq. (3) for this case are key to further analysis. That is, we are interested in $\Delta\mathbf{p} = \mathbf{p}_0 - \mathbf{p}'_0$ and $\Delta\mathbf{v} = \mathbf{v}_0 - \mathbf{v}'$. These quantities measure in some sense the "error" in the position-at-epoch and proper motion derived from the linear fit.

As noted above, the quantity to be minimized is $\sum \mathbf{e}'^2_i$, the sum of the squares of the post-linear-fit residuals, which will be denoted $X^2$:

$$
\begin{aligned}
X^2 = \sum_{i=1}^{N} \mathbf{e}_i'^2 &= \sum_{i=1}^{N} [\mathbf{p}_0 + \mathbf{v}_0 t_i + \tfrac{1}{2}\mathbf{z}t_i^2 + \mathbf{e}_i - \mathbf{p}_0' - \mathbf{v}' t_i]^2 \\
&= \sum_{i=1}^{N} [(\mathbf{p}_0 - \mathbf{p}_0') + (\mathbf{v}_0 - \mathbf{v}')t_i + \tfrac{1}{2}\mathbf{z}t_i^2 + \mathbf{e}_i]^2 \\
&= \sum_{i=1}^{N} [\Delta\mathbf{p} + \Delta\mathbf{v}t_i + \tfrac{1}{2}\mathbf{z}t_i^2 + \mathbf{e}_i]^2 \\
&= \sum_{i=1}^{N} [\Delta\mathbf{p}^2 + \Delta\mathbf{p}\cdot\Delta\mathbf{v}t_i + \tfrac{1}{2}\Delta\mathbf{p}\cdot\mathbf{z}t_i^2 + \Delta\mathbf{p}\cdot\mathbf{e}_i \\
&\qquad + \Delta\mathbf{p}\cdot\Delta\mathbf{v}t_i + \Delta\mathbf{v}^2 t_i^2 + \tfrac{1}{2}\Delta\mathbf{v}\cdot\mathbf{z}t_i^3 + \Delta\mathbf{v}\cdot\mathbf{e}_i t_i \\
&\qquad + \tfrac{1}{2}\Delta\mathbf{p}\cdot\mathbf{z}t_i^2 + \tfrac{1}{2}\Delta\mathbf{v}\cdot\mathbf{z}t_i^3 + \tfrac{1}{4}\mathbf{z}t_i^4 + \tfrac{1}{2}\mathbf{z}\cdot\mathbf{e}_i t_i^2 \\
&\qquad + \Delta\mathbf{p}\cdot\mathbf{e}_i + \Delta\mathbf{v}\cdot\mathbf{e}_i t_i + \tfrac{1}{2}\mathbf{z}\cdot\mathbf{e}_i t_i^2 + \mathbf{e}_i^2] \\
&= \sum_{i=1}^{N} [\Delta\mathbf{p}^2 + 2\Delta\mathbf{p}\cdot\Delta\mathbf{v}t_i + \Delta\mathbf{p}\cdot\mathbf{z}t_i^2 + \Delta\mathbf{v}^2 t_i^2 + \Delta\mathbf{v}\cdot\mathbf{z}t_i^3 + \tfrac{1}{4}\mathbf{z}^2 t_i^4 \\
&\qquad + 2\Delta\mathbf{p}\cdot\mathbf{e}_i + 2\Delta\mathbf{v}\cdot\mathbf{e}_i t_i + \mathbf{z}\cdot\mathbf{e}_i t_i^2 + \mathbf{e}_i^2]
\end{aligned}
\tag{5}
$$

The terms in the last row, except for $\mathbf{e}_i^2$, disappear in summation because the $\mathbf{e}_i$ are randomly oriented. We are left with

$$
X^2 = N\Delta\mathbf{p}^2 + 2\Delta\mathbf{p}\cdot\Delta\mathbf{v}\sum_{i=1}^{N} t_i + \Delta\mathbf{p}\cdot\mathbf{z}\sum_{i=1}^{N} t_i^2 + \Delta\mathbf{v}^2\sum_{i=1}^{N} t_i^2 + \Delta\mathbf{v}\cdot\mathbf{z}\sum_{i=1}^{N} t_i^3 + \tfrac{1}{4}\mathbf{z}^2\sum_{i=1}^{N} t_i^4 + \sum_{i=1}^{N}\mathbf{e}_i^2 \tag{6}
$$

Suppose $N$ is large and the observations are evenly distributed in time, at intervals $\Delta t$, from $t = 0$ to $t = T$. Then $N = T/\Delta t$ and in the limit $N \to \infty$ we have

$$
\sum_{i=1}^{N} t_i \to \frac{T^2}{2\Delta t}, \qquad \sum_{i=1}^{N} t_i^2 \to \frac{T^3}{3\Delta t}, \qquad \sum_{i=1}^{N} t_i^3 \to \frac{T^4}{4\Delta t}, \quad \text{and} \quad \sum_{i=1}^{N} t_i^4 \to \frac{T^5}{5\Delta t}
$$

Substituting the above in eq. (6), we obtain

$$
X^2 = \frac{1}{\Delta t}(\Delta\mathbf{p}^2 T + \Delta\mathbf{p}\cdot\Delta\mathbf{v}T^2 + \tfrac{1}{3}\Delta\mathbf{p}\cdot\mathbf{z}T^3 + \tfrac{1}{3}\Delta\mathbf{v}^2 T^3 + \tfrac{1}{4}\Delta\mathbf{v}\cdot\mathbf{z}T^4 + \tfrac{1}{20}\mathbf{z}^2 T^5) + \sum_{i=1}^{N}\mathbf{e}_i^2 \tag{7}
$$

If we define a coordinate system $(x, y)$ in the plane of the sky, then in that system $\Delta\mathbf{p}$, $\Delta\mathbf{v}$, and $\mathbf{z}$ are 2-vectors:

$$
\Delta\mathbf{p} = \begin{pmatrix} \Delta\mathrm{p}_x \\ \Delta\mathrm{p}_y \end{pmatrix} \qquad \Delta\mathbf{v} = \begin{pmatrix} \Delta\mathrm{v}_x \\ \Delta\mathrm{v}_y \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} \mathrm{a}_x \\ \mathrm{a}_y \end{pmatrix}
$$

The components will normally be expressed in angular units (see later discussion).

In eq. (7), the last term is a constant, as are the acceleration $\mathbf{z}$ and the time parameters $\Delta t$ and $T$. We seek values for the components of $\Delta\mathbf{p}$ and $\Delta\mathbf{v}$ that minimize $X^2$. The range of values that $X^2$ takes on can be thought of as a surface within the 4-space defined by $\Delta\mathrm{p}_x$, $\Delta\mathrm{p}_y$, $\Delta\mathrm{v}_x$, and $\Delta\mathrm{v}_y$. If we define $\nabla = (\partial/\partial\Delta\mathrm{p}_x, \partial/\partial\Delta\mathrm{p}_y, \partial/\partial\Delta\mathrm{v}_x, \partial/\partial\Delta\mathrm{v}_y)$, then the condition for minimizing $X^2$ is $\nabla X^2 = \mathbf{0}$, leading to the normal equation

$$
\frac{\partial X^2}{\partial\Delta\mathrm{p}_x} = \frac{\partial}{\partial\Delta\mathrm{p}_x}\Big[\frac{1}{\Delta t}(\Delta\mathbf{p}^2 T + \Delta\mathbf{p}\cdot\Delta\mathbf{v}T^2 + \tfrac{1}{3}\Delta\mathbf{p}\cdot\mathbf{z}T^3 + \tfrac{1}{3}\Delta\mathbf{v}^2 T^3 + \tfrac{1}{4}\Delta\mathbf{v}\cdot\mathbf{z}T^4 + \tfrac{1}{20}\mathbf{z}^2 T^5) + \sum_{i=1}^{N}\mathbf{e}_i^2\Big] = 0 \tag{8}
$$

and similarly for $\partial X^2/\partial\Delta\mathrm{p}_y$, $\partial X^2/\partial\Delta\mathrm{v}_x$, and $\partial X^2/\partial\Delta\mathrm{v}_y$. If we expand the dot products in eq. (8), (for example, $\mathbf{\Delta p}\cdot\mathbf{\Delta v} = \Delta\mathrm{p}_x\Delta\mathrm{v}_x + \Delta\mathrm{p}_y\Delta\mathrm{v}_y$), perform the indicated partial differentiations, and multiply by $\Delta t$, we obtain

$$2\Delta\mathrm{p}_x T + \Delta\mathrm{v}_x T^2 + \tfrac{1}{3}\mathrm{a}_x T^3 = 0 \tag{9$a$}$$
$$2\Delta\mathrm{p}_y T + \Delta\mathrm{v}_y T^2 + \tfrac{1}{3}\mathrm{a}_y T^3 = 0 \tag{9$b$}$$
$$\Delta\mathrm{p}_x T^2 + \tfrac{2}{3}\Delta\mathrm{v}_x T^3 + \tfrac{1}{4}\mathrm{a}_x T^4 = 0 \tag{9$c$}$$
$$\Delta\mathrm{p}_y T^2 + \tfrac{2}{3}\Delta\mathrm{v}_y T^3 + \tfrac{1}{4}\mathrm{a}_y T^4 = 0 \tag{9$d$}$$

Equations (9a) and (9c) can be solved for $\Delta\mathrm{p}_x$ and $\Delta\mathrm{v}_x$:

$$\Delta\mathrm{p}_x = \tfrac{1}{12}\mathrm{a}_x T^2 \qquad \text{and} \qquad \Delta\mathrm{v}_x = -\tfrac{1}{2}\mathrm{a}_x T$$

Equations (9b) and (9d) are identical to (9a) and (9c), respectively, except that the $y$ components of the vectors have replaced the $x$ components, so it must be true that

$$\Delta\mathrm{p}_y = \tfrac{1}{12}\mathrm{a}_y T^2 \qquad \text{and} \qquad \Delta\mathrm{v}_y = -\tfrac{1}{2}\mathrm{a}_y T$$

Obviously these results can be expressed as

$$\mathbf{\Delta p} = \tfrac{1}{12}\mathbf{z}T^2 \qquad \text{and} \qquad \mathbf{\Delta v} = -\tfrac{1}{2}\mathbf{z}T \tag{10}$$

Since the vectors $\mathbf{\Delta p}$, $\mathbf{\Delta v}$, and $\mathbf{z}$ are collinear, any dot products among them would simply equal the arithmetic product of their respective magnitudes. Also $\mathbf{e}_i^2 = e_i^2$. Therefore we can dispense with the vector notation at this point and use use the symbols $\Delta p$, $\Delta v$, and $z$ for the vector magnitudes. Equation (10) then becomes

$$\Delta p = \tfrac{1}{12}zT^2 \qquad \text{and} \qquad \Delta v = -\tfrac{1}{2}zT \tag{11}$$

The difference between the actual motion of the star and the linear model as a function of time $t$ is in the direction of $\mathbf{z}$ (+ or −). The acceleration $\mathbf{z}$ is always toward the companion, and for $\mathbf{z}$ to be considered essentially constant, the companion must be sufficiently far away that neither its direction nor distance change significantly over the short orbital arc we are considering. The magnitude of the difference between the actual motion and the linear model is

$$\delta = \Delta p + \Delta v t + \tfrac{1}{2}zt^2 = \tfrac{1}{12}zT^2 - \tfrac{1}{2}zTt + \tfrac{1}{2}zt^2 \tag{12}$$

where $z = |\mathbf{z}|$. The function $\delta$ has an extremum ($d\delta/dt$=0) at $t$=$T/2$, in the middle of the span of observations, where $\delta = -zT^2/24$. At the beginning and end of the span of observations ($t$=0 and $t$=$T$), $\delta = zT^2/12$. The total range of $\delta$ over the time interval of interest is thus $zT^2/8$. The locus of actual motion of the star on the sky crosses the least-squares-determined straight where the $\delta$ function has zeros, at $t = T/2 \pm \sqrt{T/12} = 0.211T$ and $0.789T$.

We will refer to $zT^2/24$, which is the absolute value of $\delta$ at $t = T/2$, as the *amplitude of the modelling error* from the linear approximation, designated by $h$. The total range of $\delta$ over the time span of interest is $3h$. In the following developments we will use $h$ as the metric for determining the sensitivity of the $N$ observations to the acceleration. Intuitively, it would seem that if $h$ is a few times $\langle e\rangle/\sqrt{N}$, where $\langle e\rangle$ is the mean error of a single observation (unit weight), then the observations should be at least marginally sensitive to the acceleration. More precisely, if we compute the ratio of the post-fit sum-of-squares $X^2$ for the linear and quadratic models, expressed as a function of $h$, we can use the F-test to determine the significance of the acceleration term.

To do this, we revisit eq. (7) for $X^2$, substituting in it the expressions for $\mathbf{\Delta p}$ and $\mathbf{\Delta v}$ from eq. (10). We obtain the post-fit sum-of-squares for the case of the linear fit to accelerated motion:

$$\begin{aligned} X^2 &= \frac{1}{\Delta t}\left(\tfrac{1}{144}z^2T^5 - \tfrac{1}{24}z^2T^5 + \tfrac{1}{36}z^2T^5 + \tfrac{1}{12}z^2T^5 - \tfrac{1}{8}z^2T^5 + \tfrac{1}{20}z^2T^5\right) + \sum_{i=1}^{N} e_i^2 \\ &= \frac{1}{\Delta t}\left(\tfrac{1}{720}z^2T^5\right) + \sum_{i=1}^{N} e_i^2 \end{aligned} \tag{13}$$

Here, the $e_i$ are the lengths of the 2D error vectors, that is, $e_i^2 = \mathbf{e}_i \cdot \mathbf{e}_i$. In the large-$N$ case we are considering, $N = T/\Delta t$. But $\langle e \rangle$ is defined such that $\sum_{i=1}^{N} e_i^2 = N\langle e \rangle^2$, so eq. (13) becomes

$$X^2 = N\left(\frac{1}{720}z^2T^4 + \langle e \rangle^2\right) \tag{14}$$

and the fractional increase in $X^2$ (the "extra sum of squares") due to the modeling error from the linear approximation (and assuming no other modeling errors) would be

$$\frac{\Delta X^2}{X^2} = \frac{1}{720}\frac{z^2T^4}{\langle e \rangle^2} = \left(\frac{1}{26.8}\frac{zT^2}{\langle e \rangle}\right)^2 \approx \left(\frac{h}{\langle e \rangle}\right)^2 \tag{15}$$

where $z$ and $\langle e \rangle$ are expressed in the same angular units, and $h$ is the previously defined amplitude of the modeling error, $zT^2/24$. Suppose we set $h = g\langle e \rangle/\sqrt{N}$, where $g$ is a factor to be determined; then the fractional increase in $X^2$ is $g^2/N$. The ratio $F$ that is subject to the F-test is the fractional increase in $X^2$ (= fractional increase in $\chi^2$, since all observations have a an uncertainty of $\langle e \rangle$) times the number of degrees of freedom, $\nu$, in the quadratic fit:

$$F = \frac{\Delta X^2}{X^2/\nu} = \frac{g^2}{N}\nu \tag{16}$$

If $N$ is sufficiently large that $\nu \approx N$, then $F \approx g^2$. For an F-test probability of 95%, $F$ would have to be about 4 for a wide range of degrees of freedom ($F$=4.17 for $\nu$=30, 3.92 for $\nu$=120, 3.84 for $\nu$=$\infty$). Thus we require $g = 2$, i.e., $h = 2\langle e \rangle/\sqrt{N}$. This result is not surprising, since a "two sigma" value for a model parameter determined from Gaussian-distributed data has a 95% chance of being significantly different from zero. If we wish a more stringent test, we can always set $g$=3 (probability > 99%); the value of $g$ can be adjusted according to the acceptable ratio of false positives / false negatives in the results.

## 3. Estimating the Magnitude of the Effect

This section provides an assessment of the range of orbital semimajor axes that astrometric measurements of a given accuracy and time span are sensitive to. This result would form the basis for any estimate of the fraction of stars whose observations would be affected by accelerated motion. The development given above treats the statistics of stellar loci at the weak curvature limit — those representing a very small part (a few percent) of a binary orbital period. The strong curvature limit might be plausibly defined by the point at which a reliable orbital solution becomes feasible. That point will be somewhat arbitrarily defined here as being half an orbital period, although preliminary orbital solutions are often formed from observations spanning much less time.

To assess the practical effect of unmodeled accelerations, a reasonable approach is to compare, for each candidate binary system, the amplitude of the linear-track modeling error, $h$, to some detection criterion $\epsilon$ that we are free to choose. The results of the previous section show that a reasonable choice for $\epsilon$ is $2\langle e \rangle/\sqrt{N}$, where $\langle e \rangle$ is the mean error of a single observation and $N$ is the number of observations. (More correctly, the denominator should be $\sqrt{\nu}$, where $\nu$ is the number of degrees of freedom in the quadratic solution, but we are assuming that $N$ is sufficiently large that $\nu \approx N$.)

The value of $h$ is defined for a specific time period of duration $T$, and we will use the expression $h = zT^2/24$ derived in the previous section. Although this expression somewhat underestimates the effect of orbital motion on the data for stars that traverse a significant fraction of their orbits in time $T$, it is the appropriate expression to use for the weak curvature limit.

Our task, then, is to determine over what range of conditions $h > \epsilon$. Since $h$ is proportional to the star's projected acceleration on the sky, $z$, we must first relate $z$ to the magnitude of the true acceleration of the star in 3-space, $|\mathbf{Z}|$. The true acceleration can be expressed in terms of the physical parameters of the binary system, and we can then use what is known about the distribution of these physical parameters to estimate the frequency of significant acceleration effects on astrometric data.

The magnitude of the acceleration projected onto the plane of the sky is related to the magnitude of the true 3-D acceleration vector by $z = (\sin\theta)|\mathbf{Z}|/d$, where $\theta$ is the angle between the direction of $\mathbf{Z}$ and the line of sight, and $d$ is the distance to the star. The parallax, $p$, is $1/d$, so our $h > \epsilon$ condition becomes

$$\frac{1}{24}\,(\sin\theta)\,|\mathbf{Z}|\,T^2\,p > \epsilon \tag{17}$$

The acceleration magnitude is related to the physical state of the binary system through $|\mathbf{Z}| = GM/R^2$, where $R$ is the instantaneous distance between components, and $M$ is either (a) the mass of the companion, if the star's motion is measured in an inertial system, or (b) the total mass of the system, if the motion is measured with respect to the companion. If we use units of AU, years, and solar masses, the gravitational constant $G = 4\pi^2$ and the true acceleration $|\mathbf{Z}|$ is expressed in AU year$^{-2}$. If $d$ is in parsecs and $p$ in arcseconds, then $z$ is in units of arcsec year$^{-2}$ and $h$ is in arcsec.

The distance $R$ is obviously closely related to the semimajor axis of the orbit, $a$. If we write $R = a(R/a)$ (for reasons that will become apparent shortly) our condition for detection of weak curvatures becomes

$$\frac{4\pi^2\,(\sin\theta)\,M\,T^2\,p}{24\left[a\left(\frac{R}{a}\right)\right]^2} > \epsilon \tag{18}$$

The strong curvature limit is defined by $P \geq 2T$, where $P$ is the orbital period, given by $P = a^{3/2}/\sqrt{M}$, where $M$ is the total mass of the system. For this limit we therefore have the condition $a^{3/2}/\sqrt{M} > 2T$. Rearranging the expressions for the weak and strong curvature limits, we obtain the limits on the semimajor axis that define the fraction of stars of interest:

$$(2T)^{2/3}\,M^{1/3} \;<\; a \;<\; \frac{\pi T}{\left(\frac{R}{a}\right)}\left(\frac{M\sin\theta}{6}\right)^{1/2}\left(\frac{p}{\epsilon}\right)^{1/2} \tag{19}$$

Note that $M$ on the left and right sides can have different meanings: on the left it is always the total mass of the system; on the right it is the total mass of the system only if the motion of the star is measured with respect to its companion. If the star's motion is measured with respect to an inertial frame then $M$ on the right is the mass of the companion. The expressions therefore can be used for investigating planet detection by the astrometric method, in the long-period limit, simply by setting $M$ on the right side to be the planet's presumed mass.

We have chosen $\epsilon = 2\langle e\rangle/\sqrt{N}$, which is twice the $1\sigma$ uncertainty of any angular variable derived from the observations (assuming $N \approx \nu$). Therefore, the quantity $p/\epsilon$ appearing on the right side of eq. (19) can be thought of as half the signal-to-noise ratio of the parallax: $p/\epsilon = \frac{1}{2}p/\sigma_p$. (This holds only for the case where the parallax of the star is determined using the same set of observations used for the proper motion and acceleration determination.) If we fix the $p/\sigma_p$ ratio, eq. (19) provides the semimajor axis limits for stars in a certain distance shell; for example, if we wish to consider stars with parallaxes good to 10% ($p/\sigma_p = 10$) when $\sigma_p \approx 10^{-3}$ arcsec, we will obtain the limits on $a$ for stars at a distance of approximately 100 pc. It is interesting to note in this regard that if we have a parallax-limited sample of stars uniformly distributed through space, the average parallax is only 1.5 times the minimum parallax. In such a case, and assuming that the minimum parallax is $\sigma_p$, then $p/\epsilon < 1$ for most stars in the sample, and the upper limit on $a$ will generally be quite restricted. However, most catalogs are magnitude-limited and their stellar distributions are much more concentrated toward the Sun. In such catalogs, and in special samples of nearby stars, there will be enough stars with $p/\epsilon > 1$ to provide a useful range of detectable semimajor axes.

To assess quantitatively how many stars fall between the limits defined by eq. (19), we need to know something about the distributions of $\sin\theta$, $a$, $M$, and $R/a$. Two of the four distributions, for $\sin\theta$ and $R/a$, are easy to deal with. Although the projection factor $\sin\theta$ is unknown for any specific binary system, we can assume that the direction of the vector separating the two components is randomly distributed over $4\pi$ steradians. For such a distribution, the average projection factor onto any given plane is $\pi/4 \approx 0.79$, the median projection is 0.87, and 71% of such vectors have projection factors of 0.7 or more. Therefore we can adopt $\pi/4$ as the average value of the projection factor $\sin\theta$, with some confidence that the average is also typical.

The distribution of $R/a$ values presents a similar situation. For circular orbits, $R/a$=1 at all times. For stars in eccentric orbits, $R/a$ varies between 1–$e$ and 1+$e$, where $e$ is the orbital eccentricity. However, binary stars spend more time near apastron than periastron, and over an orbital period the average value of $R/a$ is $1 + \frac{1}{2}e^2$. Since this ratio can only take on values between 1 and 1.5, we can adopt an average value of $R/a$ of 1.3 without much concern about the distribution of eccentricities.

Inserting the appropriate numerical quantities, eq. (19) becomes

$$1.59\,T^{2/3}\,M^{1/3} \;<\; a \;<\; 0.88\,T\left(M\,\frac{p}{\epsilon}\right)^{1/2} \tag{20}$$

Since $\epsilon = 2\sigma_p$, the right side (upper limit to $a$) could also be expressed as $0.62T(M\,p/\sigma_p)^{1/2}$. As we expect, the range of applicable $a$ increases with $T$. For $M = 1$ and $p/\epsilon = 1$, $7.4 \le a \le 8.7$ if $T = 10$ and $34 \le a \le 87$ for $T = 100$. These limits define the null set for $T < 6.1$ years for solar-mass binaries where the parallax is at the limit of detection but do not preclude more massive or closer systems.

If we think about the astrometric method of detecting orbital motion in general, including both orbit and acceleration solutions, the situation is this: at given distance, and for a given primary and companion mass, the astrometric sensitivity increases with semimajor axis, provided at least one full period is observed. However, once only a fraction of a period can be observed, the astrometric sensitivity decreases with increasing semimajor axis, because the observed motion becomes more linear. As the distance increases, the width of the regime of detectable semimajor axes narrows, since the observational scatter corresponds to an increasing linear scale. Thus there is a reduced range of semimajor axes that are greater than the scatter but less than that for which the orbital motion is statistically indistinguishable from a straight line.